

AMENDMENTS TO THE SPECIFICATION

Please revise the specification as follows:

[0012] Figure 1 illustrates a web site system 30 according to one embodiment of the invention. The web site system 30 includes a web server 32 that generates and serves pages of a host web site to computing devices 35 of end users. The web site provides user access to a ~~database 35~~ database 34 containing representations of items that are arranged within a plurality of item categories. A web site is one type of database access system in which the invention may be embodied; other types of database access systems, including those based on proprietary protocols, may also be used.

[0013] The items included or represented in the ~~database 35~~ database 34 may, for example, include physical products that can be purchased or rented, digital products (journal articles, news articles, music files, video files, software products, etc.) that can be purchased and/or downloaded by users, web sites represented in an index or directory, subscriptions, and other types of items that can be stored or represented in a database. Many millions of different items and many hundreds or thousands of different item categories may be represented within the item ~~database 35~~ database 34. Although a single item ~~database 35~~ database 34 is shown, the ~~database 35~~ database 34 may be implemented as a collection of distinct databases, each of which may store information about different types or categories of items.

[0015] As depicted by the query server 38 in Figure 1, the web site system 30 also includes a search engine that allows users to search the item ~~database 35~~ database 34 by entering and submitting search queries. To formulate a search query, a user types or otherwise enters a search string, which may include one or more search terms or "keywords," into a search box of a search page served by the web server 32. The search interface may also provide an option for the user to limit the search to a particular top-level browse category, or to another collection of items. In addition, the search interface may support the ability for users to conduct field-restricted searches in which search strings are entered into search boxes associated with specific database fields (author, artist, actor, subject, title, abstract, reviews, etc.).

[0016] When a user submits a search query, the web server 32 passes the search query to the query server 38, which generates and returns a list of the items that are responsive to the search query. As is conventional, the query server 38 may use a keyword index (not shown) to search the item database ~~35~~ database 34 for responsive items. In addition to obtaining the list of responsive items, the web server 32 accesses a mapping table 40 that maps specific sets of search criteria, such as specific search terms and/or search phrases, to the item categories most closely related to such search criteria. If a matching table entry is found, the web server 32 displays some or all of the related item categories on the search results page together with the responsive items (see Figures 4 and 5, discussed below). An important aspect of the invention involves the process by which the mapping table 40 is generated, as discussed below.

[0019] To detect correlations between specific search criteria and item categories, a correlation analysis component 44 periodically analyzes sets or segments of this user activity data to search for correlations. For example, the correlation component 44 may treat the search string "Java™" and the item category "books>computer languages" as being related if a large percentage of the users who searched for "Java™" within a given time period also selected an item falling with the books>computer languages category within this same time period. The analysis may also take into consideration the categories explicitly selected by users during navigation of the browse tree. For example, the correlation analysis may detect that a large percentage of the users who searched for "socks" also selected the brand-based category "apparel>Foot Locker™," and treat the two as related as a result. The correlation analysis component 44 may be implemented as a program that is executed periodically by an off-line computer system.

[0020] The use of an automated computer process to detect the search criteria/item category associations provides a number of important benefits. One such benefit is that mappings for many thousands of different sets of search criteria can be generated with very little or no human intervention. For example, mappings may be generated for each of the 5K (5 X 1024) or 10K most commonly entered search strings. Another benefit is that the mappings tend to be very accurate, as they reflect the actual browsing patterns of a large number of users. An additional

benefit is that the mappings can evolve automatically over time as new items and item categories are added to the ~~database 35~~ database 34, and as search and browsing patterns of users change.

[0021] As depicted in Figure 1, the user activity database 42 stores histories of events reported by the web server 32. The events included within the event histories preferably include both search query submissions (submissions of search criteria) and item selection actions (including item selection actions performed during category-based browsing of the ~~database 35~~ database 34). The event data recorded for each search query submission event may, for example, include the search string (search term or phrase) submitted by the user, an ID of the user or user session, an event time stamp, and if applicable, an indication of the collection(s) or type(s) of items searched. If field-restricted searching is supported, the event data may also identify the specific database field or fields that were searched (e.g., title, author, subject, etc.).

[0026] The mapping table 40 may, for example, include a separate entry for each of the M (e.g., 5K or 10K) search strings that were used the most frequently over a selected period of time. Search strings that are highly similar, such as those that are identical when capitalization, noise words ("a," "the," "an," etc.), and punctuation variations are ignored, may be treated as the same search string for purposes of generating the table 40. The mapping table 40 may be implemented using any type of data structure, or combination of data structures, that permits efficient look-up of categories. One example of a type of data structure that may be used is a hash table.

[0028] It should be noted that the item categories included in the mappings need not consist of browse categories that are ordinarily used to browse the catalog of items, but rather may include specific item attributes that may be used to form a grouping of items. For instance, a particular search string may be mapped to a particular product brand (one example of a product attribute), even though the web site's browse interface does not support browsing of the catalog by brand. Thus, for example, when a user searches for "PDA," the user may be given an option to view all products from "PalmTM" and "MindspringTM," even if the system's browse tree does not include links for either of these brands. Accordingly, any group of items that share a common

attribute (e.g., author = Clark) may be treated as an item category for purposes of implementing the invention. In this regard, a category may be represented within the mapping table 40 as a particular attribute (e.g., brand = Sony™) or attribute set (e.g., type = video and rating = G), rather than by a category name or ID.

[0039] In block 72, for each popular string, the list of categories (CAT_A, CAT_B, CAT_C ...) is sorted from highest to lowest correlation score, or equivalently, for highest to lowest degree of association with the particular search string. In addition, each such list of categories is truncated to a fixed maximum length (e.g. ten categories), so that only those categories most closely related to the particular search string are retained in each list. The result of block 72 is a set of string-to-category mappings of the form shown in Figure 1 (table 40 in exploded form). As mentioned above, the correlation score values may, but need not, be retained.

[0042] Other types of relatedness metrics may also be taken into consideration when generating the mapping table 40. For instance, the correlation data generated by analyzing the user activity data may be combined with the results of an automated content-based analysis in which the search strings are compared to item records or descriptions in the database-35 database 34. Thus, the mappings reflected in the mapping table 40 need not be based exclusively on an analysis of user activity data.

[0043] Figure 3 illustrates one example of a sequence of steps that may be performed by the web site system 30 to process a search query from a user. In block 80, the search query is executed to identify items from the database-35 database 34 that are responsive to the search criteria supplied by the user. In blocks 82 and 84, the web server 32 accesses the mapping table 40 to determine whether a table entry exists that matches the user-supplied search criteria. In embodiments in which the mappings consist of search string to category mappings, this step is performed by determining whether a table entry exists that matches the user's search string. Minor variations between search strings, such as variations in the form of a search term (e.g., singular versus plural), may be disregarded for purposes of determining whether a match exists. If no match is found, the web server generates and returns a search results page that does not

Appl. No. : 10/817,554
Filed : April 2, 2004

include category data read from the mapping table (blocks 86 and 88). In this event, a set of related categories may optionally be identified on-the-fly using an alternative method, such as a method that takes into consideration the number of items found within each category.